A response to six comments on *The Community of Advantage*

Robert Sugden

School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom

r.sugden@uea.ac.uk

Abstract

This paper responds to six contributions to a symposium on my 2018 book, *The Community of Advantage*. I defend that book's claim that most normative behavioural economics implicitly uses a psychologically ungrounded model of an 'inner rational agent'. I also defend the claim that, given the contractarian approach taken in the book, opportunity is normatively prior to welfare and to particular ingredients of well-being, such as health and perceptions of agency. I show how the Strong Interactive Opportunity Criterion proposed in the book can be extended to allow comparisons between the extent of opportunity provided by different economic regimes.

Keywords: behavioural economics; competitive market; mutual benefit; contractarianism; opportunity

JEL codes: D63, D91

I am pleased and honoured that *The Community of Advantage* (*CoA*: Sugden, 2018) is the subject of this symposium. I thank the contributors for their interest in my work, and am sorry that I have space to respond to only some of their insightful comments.

1. The vision thing

One of the pleasures of writing a book, rather than articles for journals, is having space to develop larger themes – to express what the first President Bush famously called the 'vision thing'. Geoffrey Brennan and Hartmut Kliemt recognise a 'larger purpose' in *CoA* as 'to formally restate the case for "classical liberalism" – shorn both of implausible psychological baggage, and of imposed elitist values'; the baggage they have in mind is the assumption that individuals act on rational self-interest.[1] That characterisation of the vision thing in *CoA* is broadly right, but I would express it slightly differently. What I am trying to cut away is *rationalism*.

Since the beginning of the twentieth century, when Vilfredo Pareto stripped out the original psychological underpinnings of neoclassical economics, mainstream economic theory has been based on *a priori* models of rational choice and rational belief – an approach that was consolidated by the introduction of game theory in the second half of the twentieth century. Normative economics, along with important strands of moral and political philosophy, has analysed a model world of rational agents. Rationality has been understood as mutual consistency among an agent's decisions or mental states. The starting point for *CoA* is an interpretation of the evidence accumulated by behavioural economics and psychology as showing that the processes by which human beings form judgements and make decisions are not directed towards creating that kind of consistency. It is not just that we humans can make mistakes in our reasoning, or that the decision problems we face can involve more factors than are represented in simple economic models. That we are not rational agents in the sense of the theory is not something that needs to be corrected; it is how we are. As I say in *CoA* (p. 82), we need a normative economics that does not presuppose a kind of rational human agency for which there is no known psychological foundation.

Brennan and Kliemt are right to detect a strand of anti-elitism in *CoA*. The elitism it opposes is the rationalist presumption that it is only as the attitudes of rational agents that people's preferences, beliefs and judgements warrant respect. This way of thinking allows attitudes that fail to fit a template of rationality to be dismissed as morally irrelevant errors. It is the hubris of an intellectual community that is overrating the importance of the abstract analysis of rationality and reasoning in which it specialises.

2. The inner rational agent

[1] Brennan and Kliemt's paper for this symposium, pp. 1–2.

One might have expected the development of behavioural economics over the last forty years to have led to a non-rationalistic form of normative economics. However, one of the central arguments of *CoA*, based on my earlier work with Gerardo Infante and Guilhem Lecouteux (Infante et al., 2016), is that, in its main line of development, behavioural economics continues to use the neoclassical concept of rationality as its normative standard. *CoA* characterises the common features of this approach as *behavioural welfare economics*, using the joint work of Cass Sunstein and Richard Thaler as a leading example (Sunstein and Thaler, 2003; Thaler and Sunstein, 2008). The essential idea in this work is that, because of lack of information, lack of attention, lack of self-control and limitations of cognitive ability, individuals' decisions are subject to systematic errors. The objective of behavioural welfare economics is to 'make choosers better off, *as judged by themselves*' (Thaler and Sunstein, 2008: 5, italics in original; see also Sunstein, 2014: 17–19 and Thaler, 2015: 325–326). The implicit standard of an individual's welfare is the choices she would have made in the absence of imperfections of information, attention, self-control and cognitive ability.

CoA argues that this approach assumes the existence of a mode of *latent reasoning* by which, in the absence of errors, an individual would be able to arrive at decisions that satisfy the consistency axioms of rational choice theory, and hence provide suitable data for use in welfare economics. To put this another way, the approach assumes the existence of consistent *latent preferences* – that is, preferences that would be revealed by error-free decisions. Here I am using 'revealed' only in the sense of revealed preference theory: I leave open the question of whether (as Brennan and Kliemt think,[2] but I am not sure they are right about this) behavioural welfare economics relies on a 'psychologically substantive' interpretation of latent preference. However preferences are interpreted, the implicit model of human agency is of an *inner rational agent* interacting with the world through an error-prone psychological shell. The objection is that behavioural welfare economists have not provided an account of this assumed and psychologically implausible mode of reasoning. In the absence of such an account, the whole concept of behavioural 'bias' remains ungrounded. That revealed preferences are systematically context-dependent is an observable fact for which psychological science can offer explanations, but the distinction between correct choice and error is explanatorily redundant.

Sunstein's response to this critique is one that I have heard from many behavioural economists. It is, as he says, 'brisk'.[3] In rough paraphrase: 'We all know that the criteria we use are conceptually problematic, but the cases we are dealing with are too important for us to be held back by abstract problems. It's so obvious that people are making errors that we don't need a definition of error. The effects on people's welfare are so clear-cut that we don't

---

[2] Brennan and Kliemt's paper for this symposium, p. 3.

[3] Sunstein's paper for this symposium, p. 8.

need a definition of welfare. Wake up and smell the coffee.' I have to say that I find this response frustrating. My critique addresses what behavioural welfare economists have actually written about the concepts of preference, welfare, error and bias when explaining how they reach their policy recommendations. Was that not supposed to be taken seriously?

The charge of 'fancy footwork' and 'explainawaytion' is another frustrating diversionary move.[4] When behavioural welfare economists make some supposedly obvious assumption about what an individual judges to be good for her, they often interpret any challenges as claims that that she is making some other, less plausible but still well-articulated judgement. But that is to invoke the inner rational agent model in another guise. In *CoA*, the counter-hypothesis is that context-dependence goes all the way down: people's context-dependent decisions are underpinned by context-dependent thoughts about empirical matters and context-dependent thoughts about what matters in life.

I am pleased that Malte Dold and Mario Rizzo, Johanna Thoma and Douglas Bernheim all support the inner rational agent critique, at least in relation to most existing versions of behavioural welfare economics. However, Thoma sees that critique as having force only if an agent's latent preferences are interpreted either as psychological states that exist at the time when he acts contrary to them (which she rightly sees as implausible, except in cases of self-control failure) or as descriptions of hypothetical choices. For Thoma, what is wrong with such descriptions is that they are not sufficiently subjective to satisfy the 'anti-paternalist ideal' to which (at least some) behavioural welfare economists subscribe. She maintains that the critique need not apply if we think of latent preferences as 'preferences one would arrive at through an ideal, well-informed reasoning process that starts from one's current subjective interests, and … does not alter one's fundamental subjective interests'.[5] On these grounds, and contrary to *CoA*, she endorses the choice-theoretic form of behavioural welfare economics proposed by Bernheim and Antonio Rangel and defends it against the inner rational agent critique.

Thoma's argument does not deal with what, for me, is the core of the inner rational agent critique – its unanswered questions about the psychological content of latent reasoning. Thoma goes a step further than many behavioural welfare economists do by specifying what the subjective premises for this reasoning should be, namely a person's current subjective interests. That move might work if those interests could be expressed as a set of propositions that satisfy the kind of consistency properties that rational choice theory attributes to preferences. But that supposition lacks psychological foundations too. A person's subjective sense of what matters to her may be just as context-dependent – just as responsive to

[4] Sunstein's paper for this symposium, p. 8.

[5] Thoma's paper for this symposium, p. 7.

momentary cues – as her preferences over consumption items. (Indeed, the context-dependence of the former may often explain the context-dependence of the latter.) In support of that claim, *CoA* (p. 153) quotes Daniel Kahneman's 'fortune cookie' maxim, which encapsulates a finding of experimental research into self-reported well-being: 'Nothing in life is as important as you think it is when you are thinking about it'.

3. The Bernheim–Rangel approach

In his paper, Bernheim presents a related defence of the Bernheim–Rangel approach (BR for short), which he argues is misrepresented in *CoA*. Here is my take on our disagreement. BR, as presented by Bernheim and Rangel (2009) and Bernheim (2016), aims to arrive at judgements about an individual's welfare from data about his decisions. The first step is to identify the set of decisions that are made in choice situations in which there is no impairment of the individual's 'understanding of the available options and the outcomes they imply' (Bernheim, 2016: 48). These, and only these, decisions are *welfare-relevant*, i.e., provide data that can be used in judging welfare. Where such data reveal consistent preferences ('reveal' in the sense of revealed preference theory), BR's welfare judgments correspond with those preferences: its criterion 'specializes to the standard notion of revealed preference' (Bernheim, 2016: 44). I take it that, in presenting this as a merit, Bernheim is invoking the principle that, as far as possible, welfare economics should be based on revealed preferences. However, BR also allows the possibility of inconsistencies within the welfare-relevant domain. In such cases, BR narrows down the range of possible welfare judgements by eliminating those that are compatible with none of the data. Clearly, this part of the method (which *CoA* unfairly calls 'rather unhelpful') does not invoke an inner rational agent. But what about the classification of some decisions as *not* welfare-relevant?

BR makes this classification by asking whether the decision maker 'properly understand[s] concepts central to the proper characterisation' of the relevant type of problem, and examines 'evidence concerning the processes of observation, attention, memory, forecasting, and/or learning, with the object of determining the contexts in which certain types of facts are systematically ignored or processed incorrectly' (Bernheim, 2016: 48–49). Preferences revealed in the presence of such misunderstandings are treated as welfare-irrelevant mistakes if they are directly contradicted by preferences revealed in the absence of misunderstandings. For example, suppose Joe chooses x from $\{x, y\}$ in situation A and y from $\{x, y\}$ in situation B, and that there is misunderstanding in B but not in A. If this is the only information available, BR concludes that x is better than y for Joe. Bernheim argues that there is no inner rational agent lurking behind this operation, because BR uses a concept of error that refers only to matters of 'fact or logic', and not to individuals' 'direct' judgements – judgements about what Thoma calls 'current subjective interests'.

CoA (p. 57) claims that this operation is 'similar to' the kinds of 'preference purification' used in other forms of behavioural welfare economics. I think that assessment is fair. In the example, BR observes that Joe's revealed preferences are inconsistent. On the basis of evidence about (for example) Joe's information, attention, cognitive ability and understanding of relevant concepts, BR concludes that he made a reasoning error in situation *B* – an error that led him to mischaracterise the available options or the consequences of choosing them. The welfare judgement that BR makes in situation *B* (i.e., that Joe fails to choose the option that is better for him) invokes a process of latent reasoning (i.e., the correct reasoning that would have led Joe to choose *x*) and appeals to a latent preference (i.e., the preference that would have been revealed, had Joe reasoned correctly). The latent reasoning that is being reconstructed is instrumental – starting from well-defined ends and reasoning to the best means of achieving them.

In some cases, this kind of preference purification may be unexceptionable. In an example used by Bernheim (2016: 49–50), Norma chooses between a yellow box and a red box. The yellow box contains pears and the red box contains apples, and she has been told this. Misremembering this information, she chooses the yellow box, thinking it contains bananas (situation *B*). Had she looked inside the boxes (situation *A*), she would have chosen the red box. BR infers that the red box is better for her.

But consider a more difficult case, which Bernheim (2016: 51–53) uses as his main example of the empirical application of BR. Since Bernheim was a co-researcher in this work, we can assume that he endorses its methods (Ambuehl, Bernheim and Lusandi, 2016). This project is motivated by the concern that 'people poorly understand the concept of compound interest, and that this limitation causes them to make suboptimal investment decisions' (p. 6). Ambuehl et al. report an experiment which used matched pairs of decision problems. In situation *A*, a person (say, Norma) reports her present valuation of an investment of \$*a* which will pay out \$*b* after *t* periods. In situation *B*, she reports her present valuation of an investment of \$*a* which earns a rate of return of *r* per cent per period; the compounded returns will be paid out after *t* periods. The parameters *a*, *b*, *r* and *t* are fixed so that the two investments have the same final payout. Reasonably, Ambuehl et al. treat the two investments as exactly equivalent in terms of their consequences for Norma. Reasonably too, they say: '[K]nowledge of targeted financial principles [i.e., an understanding of compound interest] is required to understand why each pair is equivalent. Someone who both possesses and fully operationalizes that knowledge will consistently ascribe the same value to both assets regardless of their preferences' (p. 6). Thus, if Norma values the investments differently, her valuations in the two situations are inconsistent. That is a case of context-dependent choice. But which (if either) decision is correct?

Ambuehl et al. reason as follows: 'Consumers who demonstrably misunderstand compound interest are presumably susceptible to error when making decisions [in situation *B*]. … [A]s long as the income stream flowing from each alternative is clearly spelled out (as it is [in situation *A*]), evidence of the same misunderstanding gives us no reason to second-guess choices'. Thus, if Norma demonstrates misunderstanding of compound interest, her valuation of the investment in situation *A* can be treated as 'the correct value' (p. 7).

But what financial economics tells us is that *terminal value* and *(internal) rate of return* are alternative metrics for summarising the cash flow from an investment. When retail investment firms describe the past performance of funds to potential customers, both metrics are commonly used. (Compare 'Since 2012, this fund has earned … per cent per year' and 'If you had invested $1000 in 2012, it would now be worth …'.) Like two different languages, neither metric is more correct than the other; the principles of compound interest are the rules *for translating between them*. Defending Ambuehl et al.'s reasoning, Bernheim argues that the underlying assumption is merely that '[Norma's] concerns are limited to the magnitudes of the immediate and delayed payoffs'.[6] If Bernheim means that she is concerned only about the cash flow, that is reasonable enough; but then it would be equally true to say that her concerns are limited to the magnitude of the rate of return.

Whichever metric has been used to describe the cash flow, Norma's problem in the experiment is to report a subjectively equivalent present value. If she were an agent in a rational choice model, this problem could be broken down into two components. The first component involves 'direct judgements' or 'subjective interests': Norma would retrieve her subjective time preference rate. The second component involves 'fact' and 'logic': she would identify the cash flow of the investment and discount it at that rate. Using this decomposition, the BR method could ask whether she had correctly identified the cash flow. But if Norma is a normal human being, untrained in finance, this representation of her reasoning is psychologically implausible. Unless her actual mental processing uses this subjective/factual decomposition, the idea that an analyst can retrieve Norma's 'factual' beliefs about the cash flow is problematic.

Notice how Ambuehl et al. sidestep this problem by enquiring, not into Norma's actual reasoning, but into her understanding of compound interest. From the fact that Norma lacks an understanding of mathematical methods that are used in theories of rational choice, BR is inferring that she is making mistakes in her everyday decisions. This is the kind of rationalistic elitism that *CoA* opposes.

4. Contractarianism

[6] Bernhiem's paper for this symposium, p.21.

CoA claims that its approach to normative economics avoids the inner rational agent critique by virtue of two features. First, it is *contractarian* in that, rather than addressing an impartial spectator or benevolent social planner, it addresses citizens as parties to potential agreements among themselves. Second, it proposes *opportunity-based* criteria that citizens can use on their own behalf when evaluating economic institutions and can license public decision makers to use when acting as their agents. Brennan and Kliemt and Sunstein are doubtful about these moves.

Brennan and Kliemt interpret my approach as claiming that 'individual choices, whatever they are, should be treated as normatively decisive', and that a person's actions 'should be honoured simply because they are hers'. They suggest that it 'rule[s] out by definition … any kind of critical assessment of preferences'.[7] Sunstein interprets my approach as requiring that voluntary agreements are 'respected'. He insists that, in considering whether this respect is deserved, 'what most matters is welfare – the kinds of lives that people [like those in his examples] are able to have'. In response to James Buchanan's 'That is all there is to it' account of the market – an account that *CoA* quotes approvingly – Sunstein says: 'Embodying voluntary exchange processes cannot be self-justifying'.[8] These remarks seem to be being made from the perspective of a third party – a moral observer who judges whether another person's actions or preferences *deserve to be* honoured or respected, and who makes this judgement by reference to what is 'normatively decisive' or what 'most matters'. But this is not the contractarian perspective as I understand it.

Writing as a contractarian, I am not trying to answer questions about what, morally speaking, *matters* – what matters, that is, not as judged by anyone in particular, but as seen from some synoptic viewpoint. I am not claiming that opportunities for voluntary transactions are the source of ultimate moral value. I am addressing individual citizens who are thinking about the value of markets to them as individuals, and about what they want governments to do for them in economic policy-making. As an analogy, suppose you are 30 years old, choosing a pension plan. What matters to you? Something that surely matters a lot is the purchasing power of the pension you can expect to receive after you retire – in other words, the extent of the opportunity set to which your pension will give you access. In judging that a larger pension would be better than a smaller one, you don't need to think much about what will matter to you at age 70 or 80 – whether that will be your ability to take a world cruise, to pay for your social care if you suffer dementia, to give your children a foothold in the housing market, or to leave a large bequest for the re-homing of stray dogs. This isn't to say that how you spend your pension doesn't matter morally, or that it is not a

---

[7] Brennan and Kliemt's paper for this symposium, pp. 5, 6 and 2 respectively.

[8] Sunstein's paper for this symposium, pp. 4 and 8.

proper subject of critical assessment – only that you don't need to make those decisions or assessments now. A financial adviser might reasonably say that, other things being equal, it is in your *interest*, or to your *advantage*, that you can expect a larger rather than smaller pension. In saying this, she would not be making any deep judgements about what matters in your life: it is not her job to make such judgements. *CoA* is using similar reasoning when it advises citizens to consider the opportunities that they derive from markets, and when it advises them that it is in their interests that public decision-makers use opportunity-based criteria. In these contexts, opportunity and voluntariness really are prior to welfare. Brennan and Kliemt are right to say that, as used in *CoA*, 'interest' and 'advantage' are not independently-defined concepts that help to explain the sense in which opportunity is valuable. They are simply *expressions* of that sense.

Am I justified in calling my approach 'contractarian'? I believe that it is contractarian in the sense that Buchanan used when describing his own work in the 1950s and 1960s. Inspired by the work of Knut Wicksell (1896/ 1958), Buchanan developed a 'voluntary exchange' theory of public goods in which the provision and finance of public goods are considered together; the underlying normative criterion is that public goods should supplied to the extent that the beneficiaries are willing to share the costs of supply, and that the beneficiaries should actually pay those shares. For both Wicksell and Buchanan, the fundamental contrast is with theories in which the optimal supply of public goods is treated as if it were something to be chosen by a welfare-maximising benevolent despot. In *The Calculus of Consent*, Buchanan and Gordon Tullock (1962) embed this analysis in a more general theory that represents collective choice as the many-person equivalent of market exchange. The guiding idea is that 'collective activity, like market activity, is a genuinely *co-operative* endeavour in which *all* parties, conceptually, stand to gain' (p. 266, emphasis in original). Buchanan and Tullock do not go into much detail about why market activity can be seen as genuinely cooperative, and their suggestion that markets work as if governed by a unanimity rule is questionable (as I explain later). *CoA* reports my attempt to flesh out Buchanan's conception of the market as a domain of mutually beneficial cooperation.

Of course, there are other ways of using the contract motif in normative theory. Sunstein is more sympathetic to the kind of social contract theory that is exemplified by John Rawls's (1971) account of agreement in an 'Original Position', which Sunstein sees as a possible foundation for a conception of welfare. For Sunstein, the merit of Rawls's approach parallels the error-correction logic of behavioural welfare economics: it 'abstract[s] from morally irrelevant factors, which may distort people's actual judgments'.[9] To put this another way, the relationship between real-world individuals and their corresponding contracting parties in the Original Position is analogous to that between a psychologically normal human

[9] Sunstein's paper for this symposium, p. 10.

being and his supposed inner rational agent. Doubts about the coherence of Rawls's account of abstract rational agency were the starting point for some of my first work at the frontier of economics and philosophy (Sugden, 1986). For me, the essential contractarianism of Rawls' work is located in his picture of a well-ordered society as a cooperative venture for mutual advantage and in his recognition of the importance of psychological stability – two themes that are central to *CoA*.

Brennan and Kliemt suggest that Buchanan and I are caught in a dilemma. As classical liberals, we are writing in favour of an extensive reliance on markets, but as contractarians we are (or should be) in favour of constitutional arrangements that are grounded in unanimous consent. In practical politics, classical-liberal economic prescriptions are a minority taste. According to Brennan and Kliemt, not acknowledging this conflict is a 'failing of the Buchanan scheme' which *CoA* inherits.[10] I disagree. In *CoA*, I am *advising* citizens that suitably regulated markets tend to work in their interests (in my financial adviser's sense of 'interest'). I am not recommending anyone to impose classical-liberal policies in the absence of democratic approval – just as, when behavioural welfare economists recommend paternalistic policies, those recommendations are made within, and subject to rejection by, a democratic process. Brennan and Kliemt's explanation of systematic differences between 'electoral' and 'market' preferences strikes me as convincing, and I recognise that my advice is addressed to citizens in their market personas.[11] That should not be surprising, given that I am writing as an economist and not as a political philosopher. But I am not claiming that either type of preference is more *authoritative* than the other.

Bernheim suggests that much of behavioural welfare economics could be restated as advice to citizens rather than as advice to social planners.[12] Up to a point, that is right: welfare economists can certainly advise citizens about the methods of policy appraisal that they should license political decision makers to use. But, as I argue in Section 3.4 of *CoA*, there is a problem if the recommended methods of policy appraisal are designed to correct citizens' errors. In such a case, the relevant errors must be acknowledged as such by the people who are liable to make them. This is why the 'as judged by themselves' clause is so important for Thaler and Sunstein, and why *CoA* scrutinises it so closely. To use this clause coherently in a contractarian argument, one must first advise citizens of an error that they are liable to make, and then try to show that, *if they accept the first part of the advice*, the recommended policy serves their interests.

---

[10] Brennan and Kliemt's paper for this symposium, p. 14.

[11] Brennan and Kliemt's paper for this symposium, p. 10.

[12] Bernheim's paper for this symposium, p. 22.

In some cases, this kind of argument makes good sense. Most consumers will probably accept that they do not read all the small print when they renew insurance contracts. Having accepted this, a consumer might reasonably approve the actions of a regulatory agency which assesses the relative importance of contract clauses and requires the most important ones to be made sufficiently salient. But consider Sunstein's example of John Jones, which relates to one of his favourite examples of justified paternalism. Sunstein (2014: 41–44, 107, 138, 140–142) supports fuel economy mandates (i.e., regulations that require car manufacturers to meet specified fleet average levels of fuel efficiency) on the grounds that these regulations lead to cost savings for consumers. The argument is that, even when exposed to regulated fuel economy labelling, car buyers fail to give due attention to information about energy costs. Thus, mandates 'give people what they really want – but fail to choose' (p. 107). But do the car buyers themselves believe they are not taking proper account of the fuel economy information they are given? Unless they do, a contractarian argument for mandates cannot get off the ground.

5. Opportunity

Any defence of the market has to come to terms with the fact that, although each market transaction requires the consent of the parties to that transaction, it does not require that of anyone else. Even in a perfectly competitive market with no technological externalities, a voluntary transaction can create *pecuniary* externalities – positive and negative effects on other people's expectations about their market opportunities. In this sense, markets do not operate as if under a unanimity rule. This is fundamental to *CoA*'s argument that 'markets cannot be fair': everyone can have a wide range of opportunities only if each person's opportunity set is liable to expand or contract according to how other people use *their* opportunities (*CoA*, pp. 190–192).

CoA recognises this fact by distinguishing between two concepts of opportunity. *Individual opportunity* is defined in terms of the extent of each individual's opportunity set. In a market economy of exchange, this is the set of alternative commodity bundles that the individual can reach by trading at market prices. Implicitly, the specification of each person's opportunity set takes account of other people's willingness to trade with her. *Interactive opportunity* is defined for each *set of* individuals in terms of what those individuals are able to achieve merely by trading with one another. Here, there is no reference to anyone's *willingness* to trade. Chapter 5 of *CoA* uses the concept of individual opportunity, and the associated Individual Opportunity Criterion, in arguing that it is in each individual's interest to have more opportunity rather than less, irrespective of whether her decisions reveal integrated preferences. Chapter 6 uses the concept of interactive opportunity, and the associated Strong Interactive Opportunity Criterion (SIOC), as a first step in a contractarian defence of the market.

CoA may not have made this distinction sufficiently clear, allowing readers to think that its analysis of opportunity is more restrictive that it really is. When Dold and Rizzo say that its analysis applies only to nested opportunity sets, they are considering only the analysis of *individual* opportunity.[13] Similarly, when Brennan and Kliemt say that my argument in favour of larger opportunity sets may not apply to cases of interactive choice, they are saying no more than *CoA* does.[14]

I am unsure what Bernheim means when he suggests that, because the 'unambiguous choice' relation P^* in the BR model is often 'not usefully discerning absent some refinement of the Welfare-Relevant Domain', my criterion – which I take to be SIOC – will probably face a similar problem.[15] In the BR model, xP^*y denotes that, for some specified individual, the welfare-relevant domain contains no decision problem in which y is chosen but x is available.[16] The lack of discernment arises from the fact that xP^*y is undefined unless the individual's choices between x and y are consistent across all choice sets and 'ancillary conditions' in the welfare-relevant domain. But this concept of consistency has no analogue in my opportunity-based model, which (except in defining competitive equilibrium as market-clearing) makes no reference to actual choices.

However, Bernheim is right to point out that SIOC is a binary criterion – one that an economy does or does not satisfy. That is indeed a limitation of my model, as presented in *CoA*. Here is a sketch of a possible extension.

Consider any *opportunity profile* $\mathbf{O} = (O_1, \ldots, O_n)$ of opportunity sets for individuals in an n-person exchange economy. Suppose that the choices that individuals actually make from these sets are jointly feasible (i.e., the market clears). For each non-empty set S of individuals, an *acquisition profile* $\mathbf{q}_S$ for S (i.e., a profile of net changes in holdings of commodities by members of S) is *feasible* if it is compatible with the total endowments of members of S. It is *allowable* if (whether feasible or not) it is capable of being chosen from the opportunity sets of members of S. It is *dominated* if there is some member i of S whose component of $\mathbf{q}_S$ is unambiguously inferior to some element of O_i (in a sense defined in *CoA*). SIOC is satisfied if, for every S, every feasible acquisition profile is either allowable or dominated. *CoA* shows that every competitive equilibrium satisfies SIOC. Let us define the *interactive opportunity set* of each S as the set of feasible acquisition profiles for S that are either allowable or dominated. Given individuals' endowments, SIOC effectively

[13] Dold and Rizzo's paper for this symposium, p. 5.

[14] Brennan and Kliemt's paper for this symposium, p. 7.

[15] Bernheim's paper for this symposium, p. 24.

[16] Bernheim's paper for this symposium, p. 12.

requires that all interactive opportunity sets are maximally large. A natural extension of this idea is to say that interactive opportunity increases if interactive opportunity sets expand.

For example, consider two exchange economies which exist side by side, with absolute barriers to trade between them. Initially, each economy is in competitive equilibrium, but with different relative prices. It is straightforward to show that if the barriers are removed and a single competitive equilibrium is established, *interactive* opportunity increases. This is true even if, as a result of price changes, some *individual* opportunity sets contract.

6. Agency

Writers who use the inner rational agent model to explain human behaviour sometimes treat that imagined agent as a person's authentic self. On this view, the psychological factors that interpose between the inner agent and the external world are alien obstacles to autonomous action. *CoA* (pp. 63–67, 83–106) challenges this rationalistic picture of human agency, arguing that a person's sense of herself as a psychological being can be an important part of her sense of identity. As I put it in a passage that Dold and Rizzo quote: Why should I think of myself as lacking in autonomy because my choices are governed by mechanisms that operate through ordinary human psychology? What else could govern them?

Dold and Rizzo sense the strong feeling behind that passage and notice an implication – that I am concerned with agency as well as opportunity. But these are concerns of different kinds. *CoA*'s concern with autonomy is its rejection of the idea that a person has autonomy by virtue of making higher-order judgements that endorse her ordinary preferences or actions. In *CoA*, autonomy is simply a person's subjective sense of being in control of her actions, and need not involve any kind of higher-order endorsement. Thus, the fact that a person acts on unstable or context-dependent preferences is not evidence of a lack of autonomy. Dold and Rizzo notice that I have distanced myself from a paper in which Buchanan (1979: 259, italics in original) claims that 'Man does not want liberty in order to maximize his utility [but] *to become the man he wants to become*'. Tying the value of liberty to meta-preferences and self-creation in this way seems a wrong move for a contractarian to make (Sugden, 2019).

Dold and Rizzo are careful not to make this move. Their conception of agency is essentially the same as mine – 'a sense or feeling of competence and autonomy on behalf of the chooser'.[17] They ground this conception in empirical psychology, citing in particular the work of Edward Deci and Richard Ryan, and express surprise that, in its treatment of agency, *CoA* does not follow its own mantra of considering human psychology '*as it really is*'.[18]

[17] Dold and Rizzo's paper for this symposium, p. 4.

[18] Dold and Rizzo's paper for this symposium, p. 10.

Although I have reservations about Deci and Ryan's picture of market relationships as not fully autonomous (Bruni and Sugden, 2013), I agree with Dold and Rizzo that a sense of competence and autonomy is an important contributor to most people's subjective well-being. I agree too that this sense tends to be promoted by well-structured environments that offer varied but surmountable challenges, feedback about the consequences of one's actions, and opportunities for the development of new tastes and aptitudes. However, I think Dold and Rizzo are misunderstanding the contractarian logic of *CoA* when they criticise it for not considering the possibility that 'individuals do not only value opportunities, but may also have an intrinsic interest in a sense of agency', and argue that agency should be considered 'as another important ingredient for welfare-improving choice situations'.[19]

In *CoA*'s contractarian perspective, opportunity is not an ingredient of welfare: it is the menu of options from which an individual can choose how to pursue whatever matters to him from his own point of view. In this sense, it is (as I said in Section 4) prior to welfare – at least, if each individual is to be treated as the judge of his own welfare. How best to sustain his own sense of agency is something that a person might consider when choosing between those options – for example, in choosing between job offers when one pays more but another offers more interesting challenges. In this respect, agency is analogous with health. It is undeniable that health is an ingredient of subjective well-being. Just as there can be good advice about how to make choices that sustain one's sense of agency, so there can be good advice about how to make choices that maintain one's health. But given what *CoA* set out to do, did it need to give such advice, either about health or agency?

Insofar as a person's sense of agency depends on decisions that he makes in situations in which no one else is involved (for example, keeping or not keeping a resolution to go for a run every morning), agency is outside the scope of a contractarian theory. Insofar as it depends on his decisions in voluntary transactions with others (for example, enrolling or not on an adult education course, choosing whether to patronise a neighbourhood grocery or a hypermarket), it falls within the scope of *CoA*'s analysis of opportunity. This is true even if the relevant voluntary transaction involves the production of a public good (for example, participating with others in an amateur football club or choral society): *CoA* follows Buchanan in considering how public goods can be supplied through mechanisms that approximate voluntary exchange (pp. 164–173).

I can see two ways in which Dold and Rizzo's concern with agency might fall within the proper scope of *CoA*'s contractarian approach but not also within the scope of opportunity, as conceived by *CoA*. One possibility is that agency is implicated in *self-acknowledged* self-control problems. A person recognises that a particular decision would

---

[19] Dold and Rizzo's paper for this symposium, pp. 8 and 4 respectively.

support his sense of agency, wishes to make that decision, but feels psychologically unable to carry it out. Anticipating this motivational conflict, he might want to impose constraints on his own future opportunities – a possibility that Buchanan (1979: 253) entertains when he discusses 'investment in becoming'. Adapting Dold and Rizzo's example of the small German town,[20] each resident might wish he could stay away from a newly opened hypermarket, fearing its effect on his preferences, but be unable to resist the temptation of its wide product range and low prices. The second possibility is that expansions of one person's opportunity have a direct effect on other people's sense of agency. (By 'direct', I mean that the effect is not transmitted through changes in the other people's opportunities, as in the case of pecuniary externalities.) For example, the mechanism might be cultural contagion: the new hypermarket shapes the preferences of the customers it attracts, and those effects then spread to other residents, reducing *their* sense of agency in a way that circumvents their choices. I must concede that *CoA* does not consider the second possibility, and has relatively little to say about the first, beyond a perhaps too casual scepticism about the frequency of self-acknowledged self-control problems.

An alternative reading of Dold and Rizzo is that their conception of contractarianism is broader than *CoA*'s. Recall Brennan and Kliemt's distinction between electoral and market preferences, and my response that *CoA*'s contractarianism is addressed to individuals in their market personas. Intuitively, it seems entirely credible that the inhabitants of a small town might have electoral preferences for maintaining its traditional shopping environment, but market preferences for shopping at a new hypermarket. Thoughts about agency – at least of the kind discussed by Dold and Rizzo – might be more salient when the context is political then when it is economic.

7. Mill and the community of advantage

The phrase 'community of advantage' is taken from a passage in John Stuart Mill's *Principles of Political Economy*. The opening pages of *CoA* explain the context of this passage and its significance for the argument that is to follow. Some readers have taken *CoA* to be presenting itself as a systematic application of Mill's philosophical ideas to the issues raised by behavioural economics. But, as Mozaffar Qizilbash points out, Mill's philosophical position is consequentialist and utilitarian, while *CoA* is contractarian, rejects the 'view from nowhere', and uses opportunity as its normative criterion. Sunstein (2014: 105–107) goes further, suggesting that Thaler and Sunstein's form of libertarian paternalism is similar in spirit to Mill's example of the unsafe bridge – a suggestion that Qizilbash rejects, I think rightly.[21] At the very least, a libertarian paternalist who wanted to use the 'Millian' label

[20] Dold and Rizzo's paper for this symposium, p. 8.

[21] Qizilbash's paper for this symposium, pp. 8–11.

would need to replace the ‘as judged by themselves’ clause with a claim to be one of Mill’s ‘competent judges’. Qizilbash concludes that, although *CoA*’s philosophy is not wholly Millian, its defence of the market has deep roots in ideas that are important to Mill.[22] That is my view too.

CoA makes no claim to be based on Mill’s philosophical premises. As *CoA* explains at the outset, Mill is used as a spokesman for a ‘liberal tradition’ of economic thought that has been an important strand in economics ever since the work of David Hume and Adam Smith. *CoA* identifies three core ideas in this tradition – that cooperation for mutual benefit is a governing principle of social life, that a competitive market is a network of mutually beneficial transactions, and that in cooperative relationships, it is for each individual to judge what counts as his benefit. Given that *CoA* is trying to avoid the rationalistic presuppositions of modern normative economics, Mill is a particularly suitable spokesman for that tradition. He was a major economist, writing at a time when economics had become an established intellectual discipline, but before it embraced rational choice theory. For the same reason, *CoA* (pp. 160–164) gives serious attention to the work of Jules Dupuit, the mid-nineteenth-century founding father of cost-benefit analysis.

The Greatest Happiness Principle of utilitarianism is a rationalistic morality of individual and collective action, but if it is to have any purchase on the real world, it must at some point engage with human psychology as it really is. Since *CoA* is trying to reconcile normative economics with human psychology, it looks at how Mill makes that connection. Mill (1861/ 1972: Chapter 3) argues that, as a morality for private individuals, utilitarianism depends on a natural human propensity for ‘social feelings’. The degree to which those feelings develop depends on the structure of society. According to Mill, they are most likely to be sustained when society is structured by relationships of voluntary cooperation for mutual benefit. Since this type of social structure is recommended by the Greatest Happiness Principle, Mill can claim that utilitarianism has psychological as well as rational consistency. Rawls’s theory of justice has an analogous logic. That theory has a rationalistic foundation in the Original Position, but it connects with human psychology through Rawls’s (1971: 16, 177, 453–462) requirement that his contracting parties look for principles that are psychologically stable. A society based on Rawls’s principles of justice is psychologically stable because it is viewed by its members as a ‘cooperative venture for mutual advantage’ (Rawls, 1971: 84) – or, as Mill would say, as a *community of advantage*. In taking that phrase as the title of my book, I am trying to express its intention to retain the moral content of the liberal tradition of economics while discarding its unnecessary rationalistic assumptions.

---

[22] Qizilbash’s paper for this symposium, p. 19.

References

Ambuehl, Sandro, Douglas Bernheim and Annamaria Lusardi (2016). The effect of financial education on the quality of decision making. NBER Working Paper 20618, revised August 2016. http://www.nber.org/papers/w20618

Bruni, Luigino and Robert Sugden (2013). Reclaiming virtue ethics for economics. *Journal of Economic Perspectives* 27: 141–164.

Bernheim, Douglas (2016). The good, the bad, and the ugly: a unified approach to behavioural welfare economics. *Journal of Benefit-Cost Analysis*. Available on CJO 2016 doi:10.1017/bca.2016.5.

Bernheim, Douglas and Antonio Rangel (2009). Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics* 124: 51–104

Buchanan, James (1979/ 1999). Natural and artifactual man. In James Buchanan, *The Collected Works of James M. Buchanan: Volume 1*. Indianapolis, IN: Liberty Fund.

Buchanan, James and Gordon Tullock (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Infante, Gerardo, Guilhem Lecouteux and Robert Sugden (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* 23: 1–25.

Mill, John Stuart (1861/ 1972). *Utilitarianism*. London: Dent.

Rawls, John (1971). *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.

Sunstein, Cass (2014). *Why Nudge?* New Haven, CT: Yale University Press.

Sunstein, Cass and Richard Thaler (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review*, 70: 1159–1202.

Sugden, Robert (1986). *The Economics of Rights, Co-operation and Welfare.* Oxford: Basil Blackwell.

Sugden, Robert (2012). The market as a cooperative endeavour. *Public Choice* 15: 365–370.

Sugden, Robert (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market* (Oxford: Oxford University Press).

Sugden, Robert (2019). What should economists do now? In Richard Wagner (ed.), *James M. Buchanan: A Theorist of Political Economy and Social Philosophy* (London: Palgrave Macmillan), pp. 13–37.

Thaler, Richard (2015). *Misbehaving: How Economics Became Behavioural*. London: Allen Lane.

Thaler, Richard and Cass Sunstein (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Wicksell, Knut (1896/ 1958). A new principle of just taxation. English translation (by James Buchanan) of German original. In Richard Musgrave and Alan Peacock (eds), *Classics in the Theory of Public Finance*, 72–118. London: Macmillan.